ARM and the change of support problem

Don Simone Daly, Krista L Gaustad, Chitra Sivaraman

Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA

Abstract

The Atmospheric Radiation Measurement Climate Research Facility collects, manages and distributes atmospheric measurements and measurement products to further research in atmospheric radiation balance and cloud feedback processes. An ARM user often needs data about points (locations or times) or blocks (areas, volumes, time periods) different from those of an existing ARM product. The ARM user has a change-of-support problem. ARM aims to help the user with this problem through value-added products, or VAPs, to change the support of ARM data to match the user's need with minimal user effort.

Keywords: change of support, objective analysis, distance weighted interpolation, spline, kriging

---

1. Introduction

The Atmospheric Radiation Measurement (ARM) Climate Research Facility collects, manages and distributes atmospheric measurements and measurement products to further research in atmospheric radiation balance and cloud feedback processes. An ARM instrument samples a specific volume at a specific rate, location and time. An ARM user, however, often needs data about points (locations or times) or blocks (areas, volumes, time periods) different from those at which an ARM variable has been observed or made available. The ARM user has a change-of-support problem. ARM aims to help the user with this problem through value-added products, or VAPs, to change the support of ARM data to match the users need with minimal user effort.

Gelfand (7) notes that the change of support problem is concerned with inference about the values of a variable at points (locations or times) or blocks (areas, volumes, time periods) different from those at which the variable has been observed. Change of support is about the prediction from points to points, points to blocks, blocks to points, and blocks to blocks. Downscaling, upscaling, disaggregation, aggregation and interpolation are about change of support.

Change of support techniques include both deterministic and stochastic approaches. From a mathematics perspective this problem is the approximation of an unknown function from exact observations of that function; the statistical view sees this as the prediction of an unknown function from uncertain observations. Change-of-support techniques have varying degrees of complexity and are based on Thiessen polygons, least squares or Lagrange polynomial interpolation, distance weighting, splines and kriging (10; 9). Objective analysis (1; 8; 5) and ordinary kriging (6) are examples of mathematical and statistical "inverse distance weighted" solutions.

Email address: `don.daly@pnnl.gov` (Don Simone Daly)

Barnes (1) developed objective analysis to predict values on a regular grid from possibly irregularly sampled points. Koch (8) improved the usability of Barnes' scheme with a meta-code to help the user parameterize, evaluate and interpret an objective analysis. Caracena (5) extended Barnes' scheme to predict gridless values using a gaussian weighting kernel with improvements in the predictions gained thru recursive application of the approximation. Barnes (2; 3; 4) reviewed developments in objective analysis, investigated the accuracy of variants of the method, and offered tuning pointers to improve predictions. These articles, and a few others, are a good place to start to understand the requirements for an ARM VAP(s) to address change of support.

Non-parametric estimation offers ways of estimating an unknown function without the specification of a parametric model of that function – a chief attraction of kernel smoothing and nonparametric estimation. Kernel smoothing provides a simple approach to modeling structure without the imposition of a possibly restrictive parametric model. Objective analysis is a very close kin of kernel smoothing, or local polynomial regression estimation, from the general class of methods for non-parametric estimation of functions. As in the case of nonparametric estimation, objective analysis is called "objective" (or "nonparametric") in that it requires no parameterized model of the unknown function. Objective analysis, like nonparametric estimation, is not parameter-free, featuring parameters such as the choice of the weighting function with its associated parameters such as span, the distribution of prediction points...

Change of support may be aided, or abetted, by the variable type (temporal/spatial, continuous/discrete, linear/non-linear scaling...) and data quality (accuracy, completeness, consistency and resolution), the correlation (spatial, temporal, other) between values, the ARM sampling protocol and the user's end use. For instance, a time-series variable may require a different estimator than a spatially distributed variable.

2. ARM data and user change of support needs

There is not one change-of-support scheme that is appropriate for all cases. Descriptions of exemplar user change-of-support requests are necessary to proceed.

3. Objective Analysis by weighted sums

Objective analysis was developed to interpolate a regularly gridded approximation from irregularly spaced observations in order to initialize numerical weather prediction models (5). Barnes (1; 2; 3; 4) proposed a multi-pass objective scheme to approximate an unknown function by an analytic function based on a Fourier integral representation. Barnes method begins with a grid determined from the spatial distribution of observations. Barne's first pass approximates the unknown function at each grid point with a distance-weighted average of the observations. Following passes replace the observations with successive residuals, i.e., the differences between the measurements and the current approximations at the observed positions, in a similar weighted average to improve the approximation. In this manner, Barne's method is a successive corrections method.

3.1. Barnes' scheme

Let f represent the unknown function and $\{x_g\}_{g=1}^{N_g}$ the grid points determined from the spatial distribution of observations. Suppose $F(x)$ represents the approximating function of f at position x. Let x_i and f_i denote the position and value of the ith observation, or sample, of f, with $\{f_i\}_{i=1}^{N_s}$

and $\{x_i\}_{i=1}^{N_s}$ representing the sets of N_s observed positions and measurements. Then, Barne's first pass approximation F_0 of f on the grid $\{x_g\}_{gi=1}^{N_g}$ is $\{F_0(x_g)\}_{g=1}^{N_g}$ with

$$F_0(x_g) = \sum_{i=1}^{N_s} \frac{w_i(x_g, x_i) f_i}{N(x_g)}$$

where

$$\begin{aligned} w_i(x_g, x_i) &= \frac{1}{4\pi\kappa} \exp\left(\frac{\|x_g - x_i\|^2}{4\kappa}\right) \\ N(x_g) &= \sum_{i=1}^{N_s} w_i \end{aligned}$$

Barne's kth pass approximation F_k of f is

$$F_k(x_g) = F_{k-1}(x_g) + \sum_{i=1}^{N_s} (f_i - F_{k-1}(x_i)) \frac{w_i(x_g, x_i)}{N(x_g)}$$

The name notwithstanding, objective analysis is controlled by a number of parameters: from κ that sets the spread of the gaussian kernel to those parameters that set the size and density of the interpolation grid to the number of successive corrections performed. Barnes (2; 3; 4) and Koch (8) address parameter selection.

3.2. Caracena's scheme

Cressman weighting function:

$$w_i(x, a) = \begin{cases} \frac{a^2 - \|x - x_i\|^2}{a^2 + \|x - x_i\|^2} & \text{if} \quad a \geq \|x - x_i\| \\ 0 & \text{if} \quad x \neq a \geq \|x - x_i\| \end{cases}$$

Caracena weighting function:

$$\begin{aligned} w_i(x, L) &= \frac{\exp\left(-\frac{\|x - x_i\|^2}{L^2}\right)}{N(x)} \\ &= \frac{\exp\left(-\left(\frac{x - x_i}{L}\right)^2\right)}{N(x)} \\ &= \frac{\exp\left(-\frac{1}{2}\left(\frac{x - x_i}{\sigma}\right)^2\right)}{N(x)} \\ &= \frac{\exp\left(-\frac{1}{2} z_i^2\right)}{N(x)} \end{aligned}$$

where

$$\begin{aligned} N(x) &= \sum_{i=1}^{N_s} \exp\left(-\frac{\|x - x_i\|^2}{L^2}\right) \\ z_i &= \sqrt{2}L \end{aligned}$$

Caracena multipass matrix scheme :

$$\begin{aligned} F_n(x) &= w^T(x, L)W^{-1}(I - (I - W)^n)f \\ &= w^T(x, L)((I + (I - W) + (I - W)^2 + ... + (I - W)^{n-1})f \end{aligned}$$

where

$$\begin{aligned} W_{ij} &= w_j(x_i, L) \\ W^{-1}(I - (I - W)^n) &= W^{-1}(I - (I - W))(I + (I - W) + (I - W)^2 + ... + (I - W)^{n-1} \\ &= (I + (I - W) + (I - W)^2 + ... + (I - W)^{n-1} \end{aligned}$$

4. Inverse Distance Weighted Averaging

4.1. Inverse Distance Weighted Interpolation

Inverse distance weighted interpolation (IDW) is a common spatial interpolation method wherein the estimated value D at location x is the weighted sum of the values $\{D_i\}$ observed at the locations $\{x_i\}$ in the neighborhood of x. IDW is a deterministic method in that the observed values are assumed known with certainty. Of note, the IDW estimate of D at location x_i is the observed value D_i. IDW assumes the similarity between the values at two locations is inversely proportional to a function of the distance between the locations. Most often, the weights follow a power or exponential distance-decay function of the distance $\text{dist}(x, x_i)$ between the locations x and x_i. The weights are normalized so that $\sum w_i = 1$. Assuming a power law distance-decay function,

$$D(x) = \begin{cases} D_i & \text{if} \quad x = x_i \\ \sum_{i=1}^{n} w_i D_i & \text{if} \quad x \neq x_i \end{cases}$$

where

$$\begin{aligned} w_i &= \frac{z_i}{\sum_{i=1}^{n} z_i} \\ z_i &= \frac{1}{\text{dist}(x, x_i)^p} \end{aligned}$$

The deterministic IDW algorithm does not account for the randomness inherent in measurements. The IDW scheme, however, can be readily extended to account for measurement uncertainty.

4.2. Inverse Distance Weighted Averaging

Inverse distance weighted averaging (IDWA) extends IDW to handle the random nature of NSCRAD scores in a inverse distance weighted calculation. Suppose the measurements Di observed at the locations xi are random draws from a $\chi^2(k)$ probability distribution. Consider the inverse distance weighted calculation

$$D(x) = \sum_{i=1}^{n} w_i D_i$$

where

$$\begin{aligned} w_i &= \frac{z_i}{\sum_{i=1}^{n} z_i} \\ z_i &= \frac{1}{(\text{dist}(x, x_i) + A)^p} \qquad \text{for} \quad A > 0 \end{aligned}$$

For values of the offset $A >> \text{dist}(x, x_i)$, z_i approximately equals $1/A^p$ for all i. Hence, w_i, for all i, approximately equals $1/n$ where n is the number of neighbors of x. In this case, the IDWA estimate $D(x)$ is simply the mean of the n neighboring D_i. When the offset $A << \text{dist}(x, x_i)$, the IDWA calculation more closely resembles IDW. The constraint that the offset $A > 0$ avoids dividing by 0 when calculating $D(x_i)$ where $\text{dist}(x_i, x_i)^p = 0$. z_i. Here, the weights w_i decrease more quickly with distance from x as the exponent p increases. For $p > 4$, almost all the weight is on the closest x_i's to x so that the IDWA estimate $D(x_i)$ about equals the NSCRAD value D_i. Similar weights may be obtain from various combinations of the offset A and exponent p. Figure **??** displays sets of weights, as curves, varying A with p fixed (panel a) and varying p with A fixed (panel b). The actual set of weights will vary with the distance from the location x, the center of the IDWA estimate. This variability may be observed in Figure **??** as the varying distances between the points on a weight curve.

5. References

[1] Stanley L Barnes. A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, 3(4):396–409, 1964.

[2] Stanley L Barnes. Applications of the barnes objective analysis scheme. part i: Effects of undersampling, wave position, and station randomness. *Journal of Atmospheric and Oceanic Technology*, 11(6):1433–1448, 1994.

[3] Stanley L Barnes. Applications of the barnes objective analysis scheme. part ii: Improving derivative estimates. *Journal of Atmospheric and Oceanic Technology*, 11(6):1449–1458, 1994.

[4] Stanley L Barnes. Applications of the barnes objective analysis scheme. part iii: Tuning for minimum error. *Journal of Atmospheric and Oceanic Technology*, 11(6):1459–1479, 1994.

[5] F Caracena. Analytic approximation of discrete field samples with weighted sums and the gridless computation of field derivatives. *Journal of the atmospheric sciences*, 44(24):3753–3768, 1987.

[6] Noel Cressie. The origins of kriging. *Mathematical Geology*, 22(3):239–252, 1990.

[7] Alan E Gelfand, Li Zhu, and Bradley P Carlin. On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1):31–45, 2001.

[8] Steven E Koch, Mary Desjardins, and Paul J Kocin. An interactive barnes objective map analysis scheme for use with satellite and conventional data. *Journal of Climate and Applied Meteorology*, 22(9):1487–1503, 1983.

[9] Jin Li and Andrew D Heap. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecological Informatics*, 6(3):228–241, 2011.

[10] Lixin Li and Peter Revesz. Interpolation methods for spatio-temporal geographic data. *Computers, Environment and Urban Systems*, 28(3):201–227, 2004.